Pergunta 1

Por responder Pontuação 1,0 Destacar pergunta

From the broadest possible perspective, data mining approaches analyze data obtained from a phenomenon to:

- a. understand past behavior or predict future behavior.
- b. predict future behavior and explain the causes of that behavior.
- c. understand or anticipate future deviations from expected behavior.
- d. Não quero responder
- e. anticipate future deviations from expected behavior and explain the causes of these deviations.

Pergunta 2

Por responder Pontuação 1,0 Destacar pergunta

In general, the most important challenge in carrying out a successful data mining project is the lack of guarantees regarding the results because:

- a. The CRISP-DM methodology does not guarantee that the business objectives are quantifiable.
- b. Não quero responder
- c. CRISP-DM methodology does not provide instruments to validate these results.
- d. The data may not contain the information needed to achieve the business objectives.
- e. the volume of data is so large that it is not possible to analyze it.

Pergunta 3

Por responder Pontuação 1,0 Destacar pergunta

Indicate in which of the following cases the reliability of the results of a poll on the vote intentions would be significantly affected. The poll is made on the basis of a sample of citizens:

- a. selected completely randomly.
- b. Não quero responder
- c. national and/or foreigners entitled to vote.
- d. with landline.
- e. with minimum age to vote, in accordance with the electoral law of Portugal.

Pergunta 4

Por responder Pontuação 1,0 Destacar pergunta

A set of data has missing values in a numeric variable, caused by data collection errors. Which of the following techniques is best suited to fill these values:
Selecione uma opção de resposta:

- a. Não quero responder
- b. the most frequent observed value.
- c. a value much higher than the highest observed value.
- d. 0.
- e. median.

Pergunta 5

Por responder Pontuação 1,0 Destacar pergunta

In a classification problem, where the objective is to predict whether a bus will be delayed in arriving at the next stop or not, the use of the data to estimate the generalization error should be:

- a. Não quero responder
- b. use a random sample of 70 percent of the data for training and the rest for testing.
- c. use a random sample of 30 percent of the data for training and the rest for testing.
- d. given day d, use data up to d-1 for training and data from d for testing.
- e. given day d, use data from d+1 for training and data up to d for testing.

Pergunta 6

Por responder Pontuação 1,0 Destacar pergunta

A prediction problem can be seen as, given a phenomenon y=f(x1, x2, ..., xN), apply a learning algorithm:

- a. to the training data, represented as (y, x1, x2, ..., xN), to assess whether f is really the function that determines the relationship between the target variable y and the independent variables (x1, x2 , ..., xN).
- b. Não quero responder
- c. to evaluate whether the value of the function f(x1, x2, ..., xN) is really y.
- d. to the training data, represented as f(y, x1, x2, ..., xN), to try to predict the value of y.
- e. to the training data, represented as (y, x1, x2, ..., xN), to try to identify f.

Pergunta 7

Por responder Pontuação 1,0 Destacar pergunta

In the k-means algorithm, as the value of the input parameter k increases, the sum of the within clusters quadratic errors:

a. increase.
b. Não quero responder
c. does not change.
d. decrease.
e. may increase or decrease.

Pergunta 8

Por responder Pontuação 1,0 Destacar pergunta

If the rule {potatos}=>{carrots} has a support of 0.5 and confidence of 1.0 then the rule {carrots}=>{lettuce} will have a support of:

a. 0.5 or higher.
b. this case is impossible.
c. Não quero responder
d. less than 0.5.
e. 0.5.

Pergunta 9

Por responder Pontuação 1,0 Destacar pergunta

One method to control overfitting in decision trees is to:

a. increase the number of levels in the tree, to minimize the probability that a leaf will have an excessive number of examples.
b. increase the number of levels in the tree, to minimize the probability that a leaf will have an insufficient number of examples.
c. reduce the number of levels in the tree, to minimize the probability that a leaf will have an excessive number of examples.
d. Não quero responder
e. reduce the number of levels in the tree, to minimize the probability that a leaf will have an insufficient number of examples.

Pergunta 10

Por responder Pontuação 1,0 Destacar pergunta

The AUC measure:

- a. Não quero responder
- b. quantifies the probability that a negative example is ranked higher than a positive example.
- c. quantifiesthe probability that a positive example is ranked higher than a negative example.
- d. represents graphically the probability that a positive example is ranked higher than a negative example as the decision threshold decreases.
- e. represents graphically the probability that a negative example is ranked higher than a positive example as the decision threshold decreases.

Pergunta 11

Por responder Pontuação 1,0 Destacar pergunta

A Neural Network is a universal approximator because:

- a. Não quero responder
- b. it can model any continuous function with only 1 node in the internal layer.
- c. it can learn any model that any other algorithm learns, including decision trees, SVM and naive Bayes, given a sufficient number of nodes.
- d. it can learn any model that any other algorithm learns, including decision trees, SVM and naive Bayes.
- e. it can model any continuous function, given a sufficient number of nodes.

Pergunta 12

Por responder Pontuação 1,0 Destacar pergunta

Metalearning can be described as:

- a. a learning approach to understand the behavior of learning algorithms.
- b. the development of super-algorithms that are immune to the No-Free-Lunch theorem.
- c. a learning approach to correct the predictions made by learning algorithms.
- d. Não quero responder
- e. the development of super-algorithms that are immune to the curse of dimensionality.